

The Apify web scraping guide for e-commerce

Why your business
needs better data
and how to get it



2022

Table of contents

1. What is web scraping?
2. Is web scraping legal?
3. What is scraped data?
4. Case study: **price monitoring**
5. Case study: **product tracking**
6. Case study: **marketing research**
7. Case study: **brand sentiment**
8. In-house vs. outsource?
10. Ready to get started?



1. What is web scraping?

Web scraping is a way of **automatically extracting information from web pages**. If you've ever copied text from a website and pasted it into a document, you were extracting that data. Web scraping uses bots to do the same thing, but much faster and more efficiently.

Web scrapers can extract **huge amounts of information in seconds**. And the data is delivered in machine-readable formats so that it can easily be used in spreadsheets, applications, and databases.

“

Every time you open a web page and copy something from it into an Excel document or into notes, that counts as scraping. Web scraping is just another name for **extracting publicly accessible information from websites**.

Ondra Urban, Apify Head of Growth



Unlock the potential of data on the web

The web is the greatest repository of knowledge and data in the history of humanity. But that information was designed to be read by human beings, not machines. Web scraping **enables you to create rules for computers to access that data in an efficient and machine-readable way**.

It is already impossible for humans to process even a fraction of the data on the web. We **need machines to read that data for us** so that we can use it in business, conservation, protecting human rights, fighting crime, and any number of projects that can benefit from the kind of data that the Internet is so good at accumulating.

To ignore the potential of web scraping is to ignore the potential of the web.

Web scraping allows you to collect **structured data**. Structured data just means that the information is easy for computers to read or add to a database.

Instead of relying on humans to read or process web pages, computers can rapidly use that data in lots of unexpected and useful ways. To illustrate the difference, imagine how long it might take you to manually copy and paste text from 100 web pages. A machine could do it in less than a second if you give it the correct instructions. It can also do it **repeatedly, tirelessly, and at any scale**.

Forget about 100 pages. A computer could deal with 1,000,000 pages in the time it would take you to open just the first few.



MYTH

"Web scraping is illegal!"

FACT

There are no laws banning web scraping on the internet.



2. Is web scraping legal?

Contrary to popular belief, there's nothing shady or illicit about web scraping. **Web scraping is legal.** That does not mean that any kind of web scraping is legal: there are boundaries. The most important boundaries are personal data and intellectual property regulations, but other factors, such as the website's terms of service, can play a role as well.

Remember to always **respect your target websites** and **use empathy to create ethical scrapers.**

“

The World Resources Institute asked Omdena to help identify economic and financial incentives for forest and landscape restoration in Latin America. One of the tasks was to get 1,000s of PDFs from over 100 government websites. The Omdena team solved the problem by using Apify to scrape Google Search Engine Results Pages (SERPs) and was able to download the entire 740,000 files in less than 14 hours.

Leo Sanchez, Head of Tech Partnerships at Omdena



Ethical web scraping as a force for good

Scraping personal data

Private personal data is always protected by law and should never be scraped. You **only have lawful access to publicly available data**. So always consider whether the information was publicly available and not behind a password authentication barrier.

Scraping copyrighted content

The scraping of copyrighted content is only permitted **for the purposes of generating information**. For example, you can scrape a web page to extract prices from it or books for natural language analysis, but you cannot scrape news articles and then republish them on your own website.

4 rules for ethical scraping

1. Scrape only public data.
2. Do not infringe on rights by scraping and using copyrighted data.
3. Do not seek to overburden the targeted website.
4. Do not use the information to steal market share.

Web scraping is legal

So, is web scraping legal? Yes, it is, but, like everything else, it can be used ethically or unethically. We believe in **web scraping as a force for good**. Apify helps rescue trafficked children, find lost dogs and even restore forests with web scraping. So it can't be all bad, can it?



MYTH

"Scrapers operate in a grey area!"

FACT

Scraping companies are regular businesses and follow rules.



3. What is scraped data?

Scraped data is any information you have retrieved from a web browser or user interface. This information could be anything from product items and price lists to photos and videos.

Scrapers **extract data in structured formats** such as XML, JSON, and Excel, so you can easily use that data in spreadsheets, databases, reports or in other apps.

Let's explore 4 different ways to use scraped data for business.

“

I don't steal personal data or trade secrets, I extract terabytes of **public data** and translate them into a language that machines can understand.

Ondra Urban, Apify Head of Growth



4 ways you can leverage scraped data

Competitor price monitoring

Online shopping makes it easy for potential customers to compare prices before paying for a product or service. With web scraping technology, you can **quickly and efficiently research and price your products and services** appropriately to maximize conversions.

Product tracking

Retailers and e-commerce enterprises use web scraping to **track listings, products, and sales data** from various online stores to see how different items and categories are performing for you and your competitors. You can then adjust your product range or use that data to target new audiences.

Market research

Scraped data can be used to audit popular e-commerce platforms and mobile apps for images, products, related keywords, and more to **stay ahead of emerging trends and consumer habits**. Learn what the market wants before your competitors.

Brand sentiment

Web scraping allows you to engage with your customers in real time so you can understand their side of the story. By **scraping customer reviews and social media chats**, you can create a data-driven strategy for understanding how to better serve your clients, or see where you need to do better.



MYTH

"Web scrapers are stealing data!"

FACT

Web scrapers only collect data that's publicly available online.



4. Price monitoring

In the e-commerce industry, price monitoring is essential if you want to stay ahead of the competition.

If you are a retailer, **price monitoring will give you greater control** over the market and will give you an indicator of how competitive your products are.

If you are a brand, price monitoring will let you know **how your products or services are positioned on the market**.

“

I really like the **flexibility and power** behind the Apify platform. It has everything you need to build customized web scrapers that **work on even the most complicated websites**. Support is excellent, with the team always on hand to answer any queries and great help site documentation.

Bryce Davies, Head of Growth at SeedLever



Why automate price monitoring?

E-commerce is a rapidly growing industry. There are now tens of millions of e-commerce websites around the world, all competing for customers. Price monitoring is critical for companies in this industry, and it is not surprising that **successful e-commerce is becoming fully automated**.

Web scraping makes it easy to simultaneously monitor millions of e-commerce websites and products in real-time. With web scraping technology you can find e-commerce products, explore social selling sites, automate research, and track competitors. This way you can streamline the **monitoring of competitor prices** and products so you can adjust your own pricing strategy and optimize your performance.

To put it simply, **price scraping** is what retailers and e-commerce businesses need to do in order to maximize their profits without pricing themselves out of the market.

Price scraping over extended periods of time can help **uncover key strategies** such as special offers during holiday seasons, know when to reduce prices and absorb losses on popular products to attract more customers, and set up pricing algorithms based on the selling history of a product.

Through price scraping, a business can create more revenue by setting up **strategic pricing techniques to attract customers** while keeping the average profit margins on your products the same.



MYTH

"Web scraping is hacking!"

FACT

Web scrapers access websites exactly the same as a human user.



CASE STUDY

- Nasdaq-listed
- E-commerce retailer
- Active in 150+ countries
- 10k+ product categories

Operating in a fiercely competitive market, a Nasdaq-listed e-commerce retailer looked to Apify to help give them a competitive edge. Selling across thousands of product categories, the retailer needed a way to collect product price data at scale to understand how they were performing against their competitors in real time.

The challenges

The retailer monitored the competition regularly but struggled to keep track of how they were performing across multiple product categories. Competitor websites were **blocking web crawlers at high volumes**, so data extraction would fail or become corrupted as data volume increased. This was a huge problem for the retailer, as their sales strategy depended on tracking the listing price of thousands products across their markets.

Our approach

With no way to purchase market intelligence data, the retailer used Apify to collect product descriptions and real-time market pricing. Apify's crawlers were **designed to behave like real users** on the target site and proxies allowed them to scale their data extraction. The extracted data was served via API so that the data team could easily integrate it into their internal reporting systems.

The results

Apify helped the client extract product data from competing websites in real time and at scale, enabling them to monitor prices, availability and other marketplace data. By collecting data on thousands of products across the globe, the retailer was **able to track how competitors were pricing their products** and identify areas of improvement. This data now fuels their data team with critical pricing data, enabling them to create powerful insights and predictions for management.



Click these ready-to-use scrapers to try for free



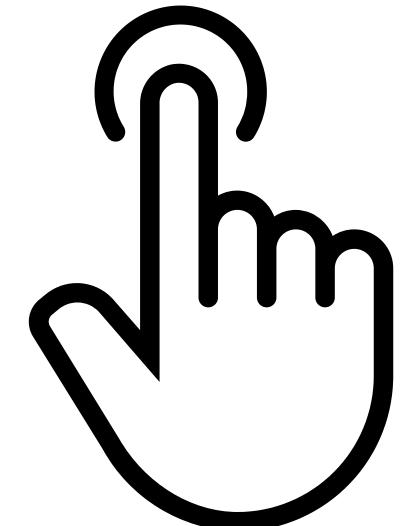
Amazon Best Sellers Scraper

Scrapes the Amazon Best Sellers categories and extracts the top 100 most popular items on Amazon. Download product name, price, URL, and thumbnail image. Works on .com, .co.uk, .de, .fr, .es, and .it domains. Download your data as HTML table, JSON, CSV, Excel, XML, and RSS feed.



eBay Scraper

Unofficial eBay API to extract data from eBay based on keywords or categories. Scrape reviews, prices, product descriptions, images, location, availability, brand, and more. Download extracted data in structured format and use it in reports, spreadsheets, databases, and applications.



5. Product tracking

Product tracking is all about extracting product details such as price, description, images, and unique product codes from websites, and **monitoring competitor performance in real time.**

In the constantly changing and competitive e-commerce market, product trends can directly affect sales strategies.

You can use product tracking to **identify new opportunities, optimize your online store, or just keep up with the competition.**

“

The Apify team was great to work with and they rolled out our scraper without any issues. The team is there whenever you contact them and their scraping solution delivers exactly the **large amounts of precise data** we need.

*Zeyn Khan, E-Commerce Specialist,
AutoMarine Online*



Why automate product tracking?

Online businesses continuously change strategies and that can make it difficult to **monitor your competitors and products**. When it comes to staying ahead of the curve, the role retail data plays in the retail ecosystem has been increasing like never before.

These days, data has become a multipurpose tool, assisting retailers big and small in specifying anything from ways to improve their product offerings to fresh ideas for creating new ones.

Web scraping makes it easy to **monitor products on millions of sites in real-time and simultaneously**. You can track listings, items, and sales data from various online stores to see how different products and categories are performing for you or your competitors.

Automated monitoring helps to **benchmark the performance of your products** at all times. It allows you to measure product performance throughout the day on a weekly, monthly, quarterly, or even an annual basis.

Scraping data makes it easy to find e-commerce products, explore social selling sites, automate research, and track competitors.

Web scraping gives you the power to automate product customization. You can adjust the range of products you offer, enter new niches, make sure that retailers are describing your products correctly, and **understand exactly how your products fit into the e-commerce landscape**.

The Apify logo, consisting of a stylized orange letter 'A' with a blue outline.

Apify advantage

Collect data from any website

Start extracting unlimited amounts of structured data right away with our ready-to-use scraping tools or work with us to solve your unique use case. Fast, accurate results you can rely on.



CASE STUDY

- Nasdaq-listed
- E-commerce retailer
- Millions of customers worldwide
- Selling across thousands of product categories

In this constantly changing and competitive market, The e-commerce retailer enlisted Apify to help collect product descriptions across multiple websites including industry giants - Google and Amazon. The data that the client needed to track included price, product descriptions, images and unique product codes.

The challenges

The retailer enlisted Apify to help collect product descriptions across multiple websites including industry giants Google and Amazon. They **needed to extract a high quantity of data** in the market. However, they found difficulties with extracting data using their previous solution. The two main sources (Google and Amazon) were updated regularly and needed constant maintenance to collect data at scale to understand how they were performing. The client's sales strategy depended on tracking thousands of new products that appear in the market and monitoring product trends.

Our approach

The client migrated from their previous AI-based solution, which was failing to accurately extract product data. Apify built **bespoke web scrapers** which automatically collect product data from Google and Amazon, monitoring competitor performance in real time. Products were matched via unique product codes to compare like products across platforms.

The results

Apify successfully built and implemented Google and Amazon scrapers which **delivered product information at scale** and with much higher accuracy and higher-quality data. Using this data, the client was able to understand which products were performing well. Apify tracked both the products as well as which competitors were selling them. Using the data collected, the client gathered insights and tracked trends related to new product development, informing company strategy and future product development.



Click these ready-to-use scrapers to try for free



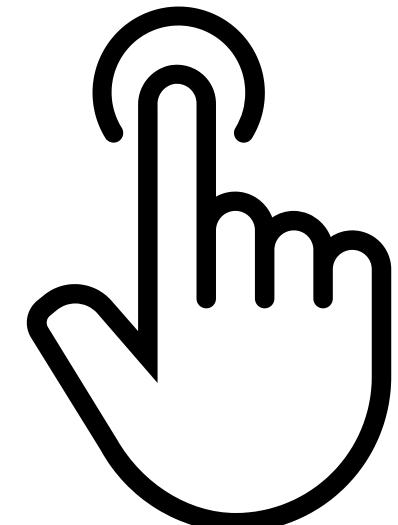
AliExpress Scraper

Scrape data from AliExpress. Extract descriptions, image, feedback, questions, prices, and all other product details. You can specify country, language, and region for shipping.



Google Shopping Scraper

Google Shopping Scraper extracts data from Google Shopping web site, in any country domain using Google SERP. It scrapes the results on the first result page.



6. Market research

Accurate and relevant information is the foundation of successful business ventures.

With web scraping, market researchers can get **invaluable insights** into market trends, competitor monitoring, research and development, and content development and analysis.

Use the **vast amount of data** on the web to gather information about existing and prospective customers, the competition, and your industry at large.

“

Apify is a key critical component of our product. Their solution fit just right with our workflow, enabling us to provide better service for our customers. The best way to reach Apify is to use their chat.

CTO of AI-powered content agency



4 market insights from web scraping

Market trends

A term that has its roots in the financial sector, 'market trends' are crucial for any industry. A market trend is anything that alters the market in which a company operates. A trend could be as simple as customer preferences, or as far-reaching as artificial intelligence technology. Such changes move fast, so keeping up-to-date with industry trends is not easy. Web scraping gives you all the information you need **in real time** to ensure you don't fall behind.

Competitor monitoring

With scraping, it's easy to monitor the competition in real time. Use this information to create a comparison site or adjust your strategy.

Research and development

A well-structured cycle of research can reduce product post-launch problems. High-quality web data opens up new possibilities for research in every aspect of the cycle. This is why companies use web scraping to uncover new trends, train AI models, and reveal new knowledge.

Content development and analysis

With web scraping you can build new online services that collect a large amount of data from selected websites, forums, and news articles, and synthesize it in a digestible format. You can send instant notifications to users about high-value updates, stay ahead of emerging trends and aggregate useful content at scale.

2

Apify advantage

Automate any online process

Scale processes, robotize tedious tasks, and speed up workflows with flexible automation software. Automation that lets you work faster and smarter than your competitors with less effort.



CASE STUDY

- **Human Coders**
- **IT training**
- **Based in Paris, France**
- **85+ online courses**

Human Coders is a group of passionate information technology professionals who offer a variety of IT courses, mostly geared towards developers who need to sharpen their skills as well as anyone looking to get started with a programming language/framework. The courses, held in French, mainly focus on the tools of the trade.

The challenges

When launching a new training course, Human Coders is always aware that it is not the only company in France covering that particular topic. Living up to its name, the company set out to make an **automated, low maintenance way** to keep an eye on its competitors' offerings. This would allow it to constantly adapt its own courses based on the data collected (duration of the course, SEO ranking of the competitors, prices, topics, etc.).

Our approach

Scraping the required information with Apify was very simple. Apify made it possible to focus on the essentials and **not risk wasting time** on sysadmin tasks. The scraping was fast enough for Human Coders to parse several thousand pages each week while keeping the resources within the bounds of the free account.

The results

Human Coders set up a series of serverless apps to run weekly on the Apify platform. These apps extract available data about all related courses offered in France. They target both their competitors as well as their own websites on a weekly basis. By using an online scraping and visualization tool, Human Coders **doesn't have to spend time maintaining servers**. Using Apify, it can instead focus on the implementation of core functionality.



Click these ready-to-use scrapers to try for free



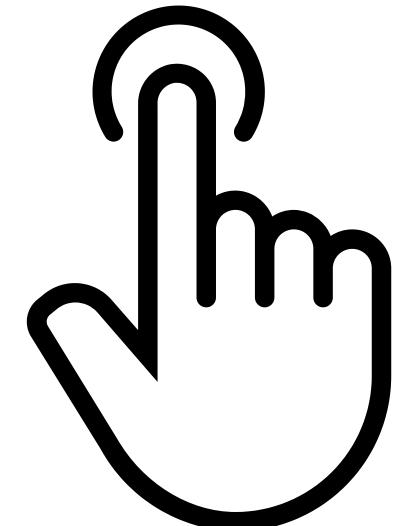
Google Search Results Scraper

Scrapes the Amazon Best Sellers categories and extracts the top 100 most popular items on Amazon. Download product name, price, URL, and thumbnail image. Works on .com, .co.uk, .de, .fr, .es, and .it domains. Download your data as HTML table, JSON, CSV, Excel, XML, and RSS feed.



Contact Details Scraper

Unofficial eBay API to extract data from eBay based on keywords or categories. Scrape reviews, prices, product descriptions, images, location, availability, brand, and more. Download extracted data in structured format and use it in reports, spreadsheets, databases, and applications.



7. Brand sentiment

Web scraping can help you understand how people feel about your brand, what they say about it, and what you need to change so that you can improve engagement and loyalty. **Social media has created new opportunities, but also increased the risks.** Negative PR or sentiment online can erode your brand if you don't monitor sentiment.

Keeping an eye on **how people are talking about your company** is now a fundamental part of marketing strategizing. Web scraping can help you by automating the data collection process.

“

Detecio uses various technologies and Apify is one of the most important and reliable. The **stability of the Apify platform and well-documented interface** enables easy integration with our internal systems.

Daniel Řezníček, Detecio co-founder



Measuring brand sentiment

Web scraping solutions allow you to **monitor thousands of product pages** and routinely extract product details, including ratings and reviews. Data can be collected regularly during the day, requiring extractions to run on schedule.

This process can be deeply integrated into your company's internal tech stack so that you can quantify brand sentiment for a truly **data-driven marketing strategy**. Brand intelligence gives you a foundation upon which to build a targeted marketing strategy in order to generate new leads and sales. Web scraping gives you a chance to do it in real-time so you can act on your customers' perceptions straight away.

There are lots of different reasons to use web scraping to track brand sentiment. Here are 3 good ones:

- **Customer insight:** understanding your customers, what they like and dislike about your product and your brand is key to building targeted marketing strategies.
- **Competition monitoring:** it's good to keep an eye on what your competitors are doing and how they are relating to their customers.
- **Real-time response:** reacting promptly to feedback shows that you are present and that you care.



Apify advantage

Integrate with any system

Export your datasets in machine-readable formats like JSON and CSV. Apify gives you APIs to let you seamlessly integrate with databases and web apps such as Zapier and Make.



CASE STUDY

- **AI-powered content agency**
- **Software marketing**
- **Based in the US & Canada**
- **200+ employees**

The client is a B2B content analysis agency with 200+ clients all over the world. Our client's solution is aimed at businesses who want to analyze user engagement on their websites. After the user behavior analysis, their mission was to figure out an algorithm to further recommend accurate, customized content of interest to these website visitors.

The challenges

The challenge before using the Apify platform was to get data from customer websites fast – thousands of web pages – and to be able to effortlessly **integrate scraped results into the database**. A smooth workflow like this would enable them to improve their understanding of the user behavior and bring maximum customer success to their work. Apify turned out to be the perfect solution to help this company advance their data extraction and scale up their analysis processes.

Our approach

It might come as a surprise, but their workflow with Apify APIs starts quite like any other user on the platform: choose a URL, edit the input parameters, run a task. Once all the necessary data is scraped, it gets effortlessly integrated into their content intelligence NoSQL database. The last and the best bit is that **Apify APIs didn't require any additional work to perform data integration**.

The results

The features that turned out to be the most useful for our client's work are the **webhooks that send the data over to the database**. Eventually, they started needing more data than the free plan was able to provide and since their experience was positive, the company upgraded to a more advanced subscription plan. Overall, our client's use of and interest in the Apify platform has been growing over time. We're now looking forward to a new level of cooperation, including developing a custom solution as well as new ways to help them scrape the web and amplify their content analysis services.



Click these ready-to-use scrapers to try for free



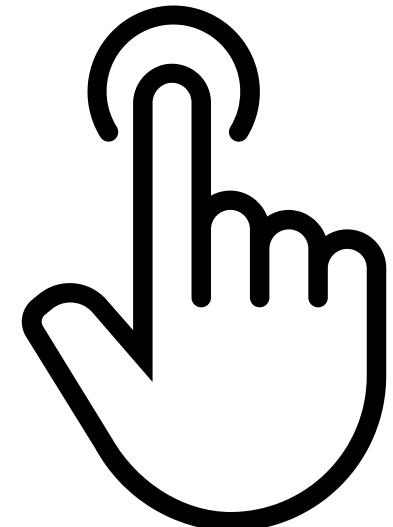
Twitter Scraper

Scrape any Twitter user profile. Creates an unofficial Twitter API to extract tweets, retweets, replies, favorites, and conversation threads with no Twitter API limits. Download your data as HTML table, JSON, CSV, Excel, XML, and use it in spreadsheets, applications, reports, and databases.



YouTube Scraper

Unofficial eBay API to extract data from eBay based on keywords or categories. Scrape reviews, prices, product descriptions, images, location, availability, brand, and more. Download extracted data in structured format and use it in reports, spreadsheets, databases, and applications.



8. In-house vs. outsource

Many companies recognize the value of big data, and they are investing in web scraping and data extraction tools – and so should you. But should you **build an engineering team dedicated to web scraping**, or should you outsource it to specialists, such as Apify?

Having an in-house team of developers dedicated to web scraping might become a burden on you and your business if you haven't made the proper considerations. Read on to find out about the risks and **how you can do both** with the Apify platform.

“

After switching to Apify, we have time back to build our core product, develop exciting features for customers. The Apify team has been flexible and innovative in accommodating our unique data collection needs and maintains the high quality and consistency of data at scale.

Bria King, Senior Product Manager at Thorn



Know the risks before you choose

Ask yourself these questions before you **decide on whether to do your web scraping in-house** or outsource:

- How much web scraping experience do I currently have in my team?
- How important is uninterrupted data extraction to my business?
- How would web scraping downtime affect my business?
- Do I have a dedicated team or am I planning to hire one?
- What is the future scope and scale of the team's web scraping activities?
- How integral is web scraping to my business?

And then make sure you're familiar with the main risks involved with doing it all in-house.

- The volume of work tends to increase with time and your team might struggle to keep up with maintenance.
- So your engineers might have little time left to analyze the data, or work on new scrapers.
- Scrapers exist across different platforms and languages.
- Your engineering team might end up being unable to satisfy your business analytics team's demand.
- Scaling web data extraction can quickly become complex and expensive.



4

Apify advantage

Scrapers that never get blocked

Smart proxy rotation of datacenter and residential proxies combine with industry-leading browser fingerprinting research to make our bots indistinguishable from humans.



Best of both with the Apify platform

There are advantages to doing your web scraping in-house, especially if you already have a team of experienced devs or an existing scraping solution.

But there are **parts of the process that you can outsource**, while still developing and maintaining your own scrapers.

The Apify platform gives you the freedom to create in-house scrapers, but **run them on a robust, scalable, and reliable scraping platform**.

You will own your code so what you scrape and how you scrape is up to you. We'll just take care of the vital infrastructure.

Alternatively, you can get a **complete end-to-end scraping solution** from Apify's world-class experts.

With a full web scraping solution from Apify, you get:

- Service-level agreement
- Your own Apify technical manager
- Dedicated development team
- Ongoing monitoring system
- Long-term maintenance
- Rich API integrations
- Data delivery direct to your systems

So **create your own web scraping solution** or **rely on us to create and maintain it**. Apify can help you with both.

5

Apify advantage

Rich developer ecosystem

Apify is built on solid open-source tools so don't worry about vendor lock-in. And you can take advantage of a thriving and skilled community of Apify Freelancers and partners.



10. Ready to get started?

Explore hundreds of ready-to-use scrapers on Apify Store or submit your project if you're ready to get started.



“

To make automation possible and strive towards the **goal of the open web as a public good** and a basic right for everyone, people need tools to extract **structured data** from the web and automate workflows on it.

Jan Curn, Apify CEO

